****

**“DESIGN A CLASSIFIER TO ASSESS THE QUALITY OF MEDICINAL PLANT USING LEARNING ALGORITHMS”**

**A DISSERTATION REPORT ON**

**Department of Computer Engineering**

**ALL INDIA SHRI SHIVAJI MEMORIAL SOCIETY’S COLLEGE OF ENGINEERING**

**KENNEDY ROAD, PUNE-411001**

**SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE**

**2024-2025**

**Under the Guidance of**

**DR. M. A. PRADHAN**

**Exam** **Seat No: 1307**

**SAYALI SHAM KAPSE**

**SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE**

**MASTER OF ENGINEERING**

**(ARTIFICIAL INTELLIGENCE AND DATA SCIENCE)**

**BY**

**A DISSERTATION REPORT ON**

**“Design a classifier to assess the quality of medicinal plant using learning algorithms”**

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE

**MASTER OF ENGINEERING**

**(ARTIFICIAL INTELLIGENCE AND DATA SCIENCE)** SUBMITTED BY

**SAYALI SHAM KAPSE Exam Seat No: 1307**

**Under the Guidance of**

**DR. M. A. PRADHAN**

****

**Department of Computer Engineering**

**ALL INDIA SHRI SHIVAJI MEMORIAL SOCIETY’S COLLEGE OF ENGINEERING**

**KENNEDY ROAD, PUNE - 411001**

**SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE**

**2024-2025**

****

**CERTIFICATE**

This is to certify that the project report entitled

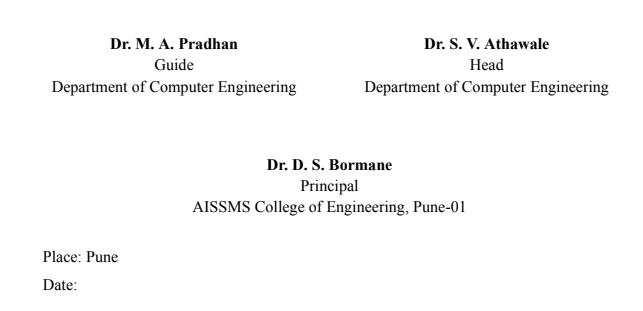
”**DESIGN A CLASSIFIER TO ASSESS THE QUALITY OF MEDICINAL PLANT USING LEARNING ALGORITHMS**”

Submitted by

**SAYALI SHAM KAPSE** Exam Seat No: 1307

is a bonafide student of this institute and the work has been carried out by her under the supervision of Dr. M. A. Pradhan and it is approved for the partial fulfilment of the requirement of Savitribai Phule Pune University, Pune for the award of the degree of Masters of Engineering (Artificial Intelligence and Data Science)

This project work has not been earlier submitted to any other Institute or University for the award of any degree or diploma.



Signature of Internal Examiner Signature of External Examiner

Place : Pune

Date:

**PROJECT APPROVAL SHEET**

A project report titles as

”**DESIGN A CLASSIFIER TO ASSESS THE QUALITY OF MEDICINAL PLANT USING LEARNING ALGORITHMS**”

is verified for its originality in documentation, problem statement, proposed work and implementation successfully completed by

**SAYALI SHAM KAPSE** Exam Seat No: 1307

at

**Department of Computer Engineering**

**ALL INDIA SHRI SHIVAJI MEMORIAL SOCIETY’S COLLEGE OF ENGINEERING**

**KENNEDY ROAD, PUNE - 411001**

**SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE**

**2024-2025**

**Dr. M. A. Pradhan**

Guide

Department of Computer Engineering

**Dr. S. V. Athawale**

Head

Department of Computer Engineering

**ACKNOWLEDGEMENT**

I would like to extend my sincere gratitude and thanks to my guide **Dr. M. A. Pradhan**, for her invaluable guidance and for giving me useful inputs and encouragement time and again, which inspired me to work harder. Due to her forethought, appreciation of the work involved and continuous imparting of useful tips, this report has been successfully completed.

I am extremely grateful to **Dr. S. V. Athawale**, Head of the Department of Computer Engineering, for his encouragement during the course of the project work.

I extend my heartfelt gratitude to the staff of Computer Engineering Department for their cooperation and support.

I take this opportunity to thank all our classmates, friends and all those who have directly or indirectly provided their overwhelming support during our project work and the development of this report.

**SAYALI SHAM KAPSE**

**ABSTRACT**

Medicinal plants like basil are widely used in traditional medicine, The Medicinal values and qualities will be reduced significantly because of plants disease. Manual inspection of these plants is time-consuming, inconsistent, and requires expert knowledge. The aim of the project is designing a classifier to assess the quality of medicinal plant using learning algorithms. We collected 666 verified basil plant images and expanded them to 2986 through augmentation techniques. Our approach involved pre-processing steps like image resizing, noise reduction, and normalization. We tested various machine learning algorithms, with CNN showing the best performance (89.22% accuracy), followed by SVM (85.89%) and Random Forest (82.66%). This system could help herbalists, and pharmaceutical companies quickly assess medicinal plant quality, and ensure higher standards in herbal medicine production.

**Keywords: Medicinal plant quality assessment, herbal medicine quality control, feature extraction, traditional and modern evaluation techniques, sustainable quality control practices, artificial intelligence and image processing, machine learning**

|  |  |
| --- | --- |
| **ABBREVIATION** | **ILLUSTRATION** |
| SVM | Support Vector Machine |
| KNN | K-Nearest Neighbour |
| LDA | Linear Discriminant Analysis |
| GBN | Gradient Boosting |
| CNN | Convolutional Neural Network |

**LIST OF ABBREVIATIONS**

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Sr.no** | **Content** | **Page no** |
| 1 | Level 0 DFD | 21 |
| 2 | Level 1 DFD | 21 |
| 3 | System Architecture of Proposed classifier | 23 |
| 4 | Working Methodology | 24 |
| 5 | Model selection strategy | 25 |
| 6 | Healthy Leaf Basil | 25 |
| 7 | Downy Mildew Disease | 26 |
| 8 | Fusarium Wilt Disease | 26 |
| 9 | Gray Mold Disease | 26 |
| 10 | Septoria Leaf Spot Disease | 26 |
| 11 | Resized Image | 27 |
| 12 | Contrast Enhancement | 27 |
| 13 | Noise Reduction | 27 |
| 14 | White balance correction | 28 |
| 15 | Concluding results after applying all image pre-processing techniques | 28 |
| 16 | Data Augmentation of sample images | 29 |
| 17 | Architecture of CNN | 31 |
| 18 | Class Diagram | 31 |
| 19 | Sequence Diagram | 32 |
| 20 | Deployment Diagram | 33 |
| 21 | Home Page | 38 |
| 22 | Input Section | 38 |
| 23 | Prediction as Healthy | 39 |
| 24 | Prediction as diseased basil -1 | 39 |
| 25 | Prediction as diseased basil -2 | 40 |
| 26 | Prediction as diseased basil -3 | 40 |
| 27 | Prediction as diseased basil -4 | 41 |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **Sr.no** | **Content** | **Page no** |
| 1 | Plan of Work | 13 |
| 2 | List of Probability Matrix | 16 |
| 3 | System Implementation Plan | 22 |
| 4 | Test Case - 01 | 36 |
| 5 | Test Case - 02 | 36 |
| 6 | Test Case - 03 | 37 |
| 7 | Test Case - 04 | 37 |

**TABLE OF CONTENT**

|  |  |
| --- | --- |
| LIST OF ABBREVIATIONS | i |
| LIST OF FIGURES | ii |
| LIST OF TABLES | iii |

|  |  |  |
| --- | --- | --- |
| **Sr.no** | **Content** | **Page no** |
| 1 | INTRODUCTION | 6 |
|  | 1.1 OVERVIEW | 6 |
|  | 1.2 MOTIVATION | 7 |
|  | 1.3 PROBLEMDEFINITION | 7 |
| 2 | LITERATURE SURVEY | 8 |
|  | 2.1 EXISTINGWORK | 8 |
|  | 2.2 PROPOSED WORK | 12 |
| 3 | DISSERTATION PLAN | 13 |
|  | 3.1 TIMELINE OF PROJECT | 13 |
|  | 3.2 FEASIBILITY STUDY | 13 |
|  | 3.2.1 TECHNICAL FEASIBILITY | 13 |
|  | 3.3.2 ECONOMICALFEASIBILITY | 14 |
|  | 3.3.3 OPERATIONAL FEASIBILITY | 14 |
|  | 3.3.4 TIME FEASIBILITY | 15 |
|  | 3.3 RISK ANALYSIS | 15 |
| 4 | SOFTWARE REQUIREMENT SPECIFICATION | 17 |
|  | 4.1 INTRODUCTION | 17 |
|  | 4.2 PROJECT SCOPE | 17 |
|  | 4.3 EXTERNAL INTERFACE REQUIREMENT | 17 |
|  | 4.3.1 USER INTERFACE | 18 |
|  | 4.3.2 HARDWARE INTERFACE | 18 |
|  | 4.3.3 SOFTWARE INTERFACE | 18 |
|  | 4.3.4 COMMUNICATIONINTERFACE | 18 |
|  | 4.4 NON- FUNCTIONAL REQUIREMENTS | 18 |
|  | 4.5 OTHER REQUIREMENTS | 19 |
|  | 4.5.1 DATASET RREQUIREMENT | 19 |
|  | 4.5.2 HARDWARE REQUIREMENTS | 20 |
|  | 4.5.3 Software Requirements | 20 |
|  | 4.6 ANALYSIS MODEL | 20 |
|  | 4.6.1 DATA Flow Diagram | 20 |
|  | 4.7 SYSTEM REQUIREMENT PLAN | 22 |
| 5 | SYSTEM DESIGN | 23 |
|  | 5.1 SYSTEM ARCHITECTURE | 23 |
|  | 5.1.1 DATASET COLLECTION AND PREPRATION | 25 |
|  | 5.1.2 DATA PREPROCESSING | 26 |
|  | 5.1.3 DATASET SPLIT | 28 |
|  | 5.1.4 DATA AUGMENTATION | 28 |
|  | 5.1.5 FEATURE EXTRACTION AND PCA | 29 |
|  | 5.1.6 MODEL SELECTION AND PERFORMANCE EVALUATION | 29 |
|  | 5.2 UML DIAGRAMS | 31 |
| 6 | TECHNICAL SPECIFICATIONS | 34 |
|  | 6.1 ADVANTAGES | 34 |
|  | 6.2 APPLICATIONS | 35 |
| 7 | TESTING | 36 |
|  | 7.1 SYSTEM TESTING | 36 |
|  | 7.2 INTEGRATION TESTING | 36 |
|  | 7.3 UNIT TESTING | 36 |
| 8 | RESULTS AND DISSCUSSIONS | 38 |
|  | 8.1 RESULTS | 38 |
|  | 8.2 DISSCUSSIONS | 42 |
| 9 | CONCLUSION & FUTURE WORK | 44 |
| 10 | REFERENCES | 45 |

Chapter 1

**INTRODUCTION**

**1.1 Overview**

Medicinal plants have played an essential role in human health for thousands of years, serving as the foundation for many traditional healing systems and modern pharmaceuticals. Today, with increasing awareness of natural and holistic health practices, these plants continue to hold significant therapeutic value. Ocimum basilicum (commonly known as Basil or Tulsi) is one such herb that is widely recognized for its medicinal properties, including its ability to alleviate digestive issues, reduce inflammation, and combat microbial infections. Especially in countries like India, Tulsi is not only used for medical purposes but is also deeply rooted in cultural and spiritual practices.

With the global herbal medicine market projected to grow from $216.40 billion in 2023 to $437 billion by 2032, there is a growing need for reliable methods to ensure the quality and safety of medicinal plants. Unfortunately, several factors such as climate conditions, soil quality, harvesting techniques, pest infestations, and plant diseases can significantly affect the chemical composition and medicinal efficacy of these plants. In particular, the presence of plant diseases can lead to a reduction in the active phytochemical components, rendering the plant less effective or even harmful for therapeutic use.

Traditional quality assessment methods largely rely on manual inspection by experts or laboratory-based chemical testing. These methods are often labour-intensive, expensive, time-consuming, and susceptible to human error or bias. Moreover, the visual identification of plant diseases by farmers or botanists may not be sufficient for early-stage disease detection or for assessing large-scale plantations efficiently.

To overcome these challenges, the integration of modern technologies such as image processing, computer vision, and machine learning has emerged as a revolutionary solution in the field of agriculture and plant health monitoring. These technologies offer the potential to automate the identification and classification of plant diseases through non-invasive, fast, and accurate analysis of plant images. In recent years, machine learning algorithms have demonstrated excellent capabilities in classifying plant health conditions by learning from labelled datasets of healthy and diseased plant images.

However, most of the existing research and applications have primarily focused on food crops such as rice, wheat, maize, and tomatoes, while medicinal plants despite their growing importance—remain relatively underexplored in this domain. This gap is significant, as the medicinal value of plants depends not only on visual features but also on their biochemical and therapeutic integrity, which can be compromised due to disease.

In this research project, we aim to bridge this gap by designing a machine learning-based classifier that can assess the quality of medicinal plants specifically Basil by analysing their leaf images. Our system is trained to detect and classify four common diseases that affect Basil: Downy Mildew, Fusarium Wilt, Gray Mold, and Septoria Leaf Spot. Each of these diseases can reduce the plant’s medicinal potency and render it unfit for consumption or therapeutic use if not identified early.

The classifier uses deep learning techniques to extract relevant features from leaf images, such as color patterns, texture, and shape irregularities, to distinguish between healthy and diseased plants. By automating this process, the system not only reduces dependency on expert evaluations but also enables faster and more consistent monitoring of plant health across large-scale farms or nurseries.

Moreover, the proposed system contributes toward the broader goal of promoting sustainable agriculture and improving the quality control of herbal medicine production. With accurate and timely detection of disease-affected plants, farmers can take preventive actions early, thereby preserving plant quality and minimizing crop losses. Additionally, pharmaceutical companies and herbal medicine manufacturers can use such systems to ensure the sourcing of high-quality raw materials, leading to safer and more effective herbal products.

In summary, this seminar report presents a complete workflow for building a classifier to assess the quality of medicinal plants using learning algorithms. The forthcoming sections will detail the literature review, methodology, data collection and preprocessing, model training and evaluation, followed by discussions on the results and future directions for improvement and real-world application.

**1.2 Motivation**

The traditional methods of assessing medicinal plant quality are slow, expensive, and time-consuming. This project is motivated by the vision to combine modern AI techniques with age-old herbal knowledge to build a reliable, fast, and accessible plant quality assessment system.

**1.3 Problem statement**

To design and develop a classifier using learning algorithms that can assess the quality of Basil leaves by identifying whether the plant is healthy or affected by disease.

7

Chapter 2

**LITERATURE SURVEY**

**2.1 Existing work:**

Evaluating the quality of medicinal plants is essential to ensure their therapeutic effectiveness and safety. Traditional methods typically depend on chemical analysis, which can be both time-consuming and often destructive to the samples. However, recent progress in machine learning (ML) and computer vision has introduced opportunities for non-destructive, fast, and precise quality assessment. This literature review examines various studies and approaches that use ML algorithms for assessing medicinal plant quality, focusing on the strengths and limitations found in current research.

T. Shen et al. (2021)[1] identified the impacts of climate change and habitat suitability on the distribution and quality of Gentiana rescans, a medicinal plant. They employed a multiple information integration strategy, combining field investigations using GPS data with chemical analyses such as high-performance liquid chromatography (HPLC) and Fourier transform infrared spectrometry (FT-MIR). This approach demonstrated the potential of integrating diverse data sources for a comprehensive quality assessment. Min He et al. (2021)[2] investigated the use of chemometric tools in chromatography-mass spectrometry to identify "material basis-Quality markers" in Chinese herbal medicines. Their review covered various aspects, including design modelling, optimization, calibration, resolution of co-eluted peaks, and fingerprint-efficacy modelling. Despite these advancements, the study noted that relying solely on chromatography-related technology does not fully disclose the material basis or quality markers in traditional Chinese medicines.

Nisar Hussain et al. (2019)[3] reviewed classical and emerging non-destructive technologies for the safety and quality evaluation of cereals. Classical methods such as HPLC and gas chromatography, while effective, have limitations due to their destructive nature. In contrast, emerging non-destructive methodologies like hyperspectral imaging, fluorescence spectroscopy, and near-infrared spectroscopy show promise for online monitoring and evaluation. These techniques could be adapted for medicinal plant assessments to provide rapid and non-destructive quality analysis.

8

Nikita Modupalli et al. (2021)[4] demonstrated the use of NIR, FTIR, and Raman spectroscopy in detecting adulterants in spices. They employed chemometrics and data analytics such as Principal Component Analysis (PCA), Partial Least Squares Discriminant Analysis (PLS-DA), and Partial Least Squares (PLS) for multivariate analysis. These methods have significant potential for the authentication and quality assessment of medicinal plants. Jingjing Wang, Quansheng Chen, Tarun Belwal, Xingyu Lin, Zisheng Luo, et al. (2021)[5] discussed the application of chemometric algorithms in foodstuff analysis using Raman spectroscopy and Surface-Enhanced Raman Scattering (SERS). The study provided a comprehensive overview of spectral pre-processing, qualitative algorithms, variable selection methods, and quantitative algorithms, indicating the growing importance of ML in quality assessment. Puneet Mishra, Jean Michel Roger, Douglas N. Rutledge, Ernst Woltering, et al. (2020)[6] introduced the SPORT (Sequential Pre-processing through Orthogonalization) approach to improve the predictive power of multivariate models based on NIR spectra. This method provides complementary information by combining multiple preprocessing techniques, enhancing the model's performance for food materials analysis, which could be adapted for medicinal plants.

YunLi, Yao Shen, Chang-liang Yao, and De-an Guo (2020)[7] reviewed recent analytical techniques used to generate chemical fingerprints for herbal medicines (HM). These techniques include chromatography, vibrational spectroscopy, nuclear magnetic resonance spectroscopy, and mass spectrometry. The study emphasized the use of chemometrics methods for data analysis but noted the absence of chemical fingerprint based chemometrics analysis for conventional HM quality assessment.

Luming Qi, Furong Zhong, Yang Chen, Shengnan Mao, Zhuyun Yan, Yuntong Ma, et al. (2020)[8] developed a machine learning model using Random Forest (RF) algorithms to trace the geographical origins of embolic medicines. Their integrated spectroscopic strategy, which included spectral pre-treatment, outlier diagnosis, feature selection, data fusion, and ML algorithms, was effective in controlling quality based on geographical origin. Khalid Tahria, Carlo Tiebeb, Nezha El Baric, Thomas Hübertb, Benachir Bouchikhia, et al. (2017)[9] employed electronic sensing systems coupled with multivariate analysis to detect adulteration in cumin. This technique successfully discriminated cumin samples from different geographical origins and quantified adulteration percentages. These methodologies have the potential to be adapted for evaluating the quality of medicinal plants.

9

Clara Pérez-Ràfols, et al. (2023)[10] used a data fusion approach combining UV-vis spectroscopic and chromatographic data to achieve higher cauterization ability in the authentication of soothing herbs. This study demonstrated the benefits of data fusion in the discrimination of complex systems, suggesting its application in the quality assessment of medicinal plants. Jianqing Zhang, Cuicui Wang, et al. (2022)[11] employed integrated untargeted metabolite profiling, cross-validation, absolute quantification, and a support vector machine model for classifying and predicting herbal medicines from multiple botanical origins. This method, exemplified by Rhizoma Alismatis, combined ultra-high-performance liquid chromatography with LTQ-Orbitrap mass spectrometry for metabolite profiling, providing a robust framework for herbal medicine authentication. Silky Sachar & Anuj Kumar (2022)[12] employed deep learning models, specifically convolutional neural networks (CNNs), for the identification of medicinal leaves. They used transfer learning to pre-train models such as MobileNetV2, InceptionV3, and ResNet50. Despite the limited dataset, their approach demonstrated the potential of deep learning in plant classification tasks. S. S. Roopashree, J. Anitha, et al. (2021)[13] developed "Deep Herb," a vision-based system using Exception features for classifying medicinal diseases. They utilized artificial neural networks (ANN) and support vector machines (SVM), but faced challenges with multi-class classification due to data imbalance.

Frimpong Twum et al. (2022)[14] used Log Gabor Filters for the textural analysis of medicinal plants, employing k-means clustering and supervised classifiers such as SVM and decision trees. Despite the innovative approach, they reported low detection accuracy. Jibi G. Thanikkal, Ashwani Kumar Dubey, and M. T. Thomas (2020)[15] introduced a Unique Shape Descriptor Algorithm for medicinal plant identification using an abridged image database. Although the algorithm proved effective, it was time-consuming and required extensive data. Congcong Wang, Xiaobo Zhang, et al. (2022)[16] utilized Random Forest algorithms to classify medicinal plants Astragalus Mongholicus Bunge and Sophora Flavescens Aiton using data from GaoFen-6 and multitemporal Sentinel-2. They addressed the issue of overfitting with heterogeneous datasets by selecting features based on their global separability index.

Kalananthni Pushpanathan, Marsyita Hanafi, et al. (2020)[17] reviewed machine learning classifiers for medicinal plant recognition, categorizing them based on their performance in classifying leaf images. They highlighted the computational challenges associated with redundant feature selection. Jing Wei Tan, Siow Wee Chang et al. (2020)[18] proposed a CNN-based method. The study found low accuracy for ANN models but demonstrated the potential of CNNs. L. C., et al. (2020)[19] discussed the biodiversity and conservation of medicinal and aromatic plants, noting that around 8000 medicinal plant species are used by different communities in India across various ecosystems. The study emphasized the need to encourage the multiplication and cultivation of these plants to ensure their availability and sustainable use.

10

Jana Wäldchen, Patrick Mäder, et al. (2020)[20] conducted a systematic literature review on plant species identification using computer vision techniques. They examined features such as shape, texture, color, margin, and vein structure, highlighting the high computational time required for these methods. Umair Ahmad, Sidra Ashiq, Gran Badshah, and Ali Haider (2022)[21] discussed the extraction of plant leaf features using deep learning. They applied convolutional neural networks (CNN) to analyse camera images or a dataset of images, noting challenges related to limited datasets and the assembly of machine learning models. Esther almerón-Manzano, Jose Antonio Garrido-Cardenas, et al. (2021)[22] examined global research trends on medicinal plants, noting a shift in focus from the cultivation or domestication of plant species to the search for new medicines or active compounds. This trend highlights the evolving priorities in medicinal plant research.

Halimatu Sadiyah, et al. (2017)[23] applied convolutional neural networks (CNN) in precision agriculture for plant image recognition and classification. They collected a database of images using remote sensing techniques and developed models to determine appropriate treatment plans for different crop types and regions, optimizing production on a maize plantation. T. Meenakshi et al. (2023) used logistic regression algorithms for the detection of diseases in medicinal plant leaves. Their study highlighted the significant impact of diseases on crop quality and yield, underscoring the importance of effective disease detection methods in maintaining the quality of medicinal plants.

Neeraj Kumar, Peter N., et al. (2012) developed "Leafsnap," a computer vision system for automatic plant species identification. The system covers 184 tree species of the Northeastern United States, employing classification and segmentation techniques, though it remains challenging for non-experts. Despite significant advancements, several gaps remain. Many studies, such as those by Sachar & Kumar (2022)[12] and Ahmad et al. (2022)[21], emphasize the need for larger and more diverse datasets to improve the robustness of ML models. Additionally, Pushpanathan et al. (2020)[17] highlighted the importance of effective feature selection methods to reduce computational costs and enhance model performance. Future research should focus on integrating chemical, spectral, and ML approaches to develop comprehensive, real-time, and non-destructive quality assessment systems.

11

**2.2 Proposed Work:**

**● Aim:**

The aim of the project is designing a classifier to assess the quality of medicinal plant using learning algorithms

**● Objectives:**

* Extensive study of state of art literature on quality assessment of medicinal plant.
* To identify and gather data relevant to medicinal plant quality
* To perform data pre-processing.
* Design classifiers using learning algorithms.
* To select the best classifier and optimize it.
* To validate the efficiency of proposed classifier.

12

Chapter 3

**DISSERTATION PLAN**

3.1 TIMELINE OF PROJECT



Table 3.1: Plan of Work

**3.2 FEASIBILITY STUDY**

**3.2.1 TECHNICAL FEASIBILITY**

A technical feasibility study is an evaluation conducted to determine the practicality and viability of implementing a specific project or technology from a technical perspective. It assesses whether the project can be successfully implemented using the available resources, technology, and infrastructure.

The proposed project – “Design a Classifier to Assess the Quality of Medicinal Plants using Learning Algorithms” – appears to be technically feasible. Deep learning techniques, particularly Convolutional Neural Networks (CNNs), have shown strong performance in image classification tasks, including plant disease detection. These models can automatically learn complex patterns from raw images, eliminating the need for manual feature extraction.

Careful preprocessing such as resizing, noise reduction, and normalization has helped maintain model consistency and performance.

13

To ensure technical success, a detailed study was conducted, including evaluating dataset quality, preprocessing techniques, compatibility with available tools, and performance of various classification models. Comparative analysis confirmed CNN as the best-performing model with 89.22% accuracy.

The proposed system uses the following technologies:

* **Deep Learning Framework**: TensorFlow and Keras (for CNN model training and evaluation)
* **Programming Language**: Python (due to its flexibility and vast AI/ML ecosystem)
* **Image Processing Libraries**: OpenCV (for preprocessing such as resizing, noise filtering, and augmentation)
* **Integrated Development Environment (IDE)**: Visual Studio Code and Jupyter Notebook (for development and debugging)

Given the availability of the required dataset, tools, and infrastructure, and based on model performance, the system is considered **technically feasible and implementable in real-world scenarios**.

**3.2.2 ECONOMICAL FEASIBILITY**

The economical feasibility of this project depends on several key factors. While there may be initial costs associated with hardware, software tools, dataset acquisition, and system development, the potential long-term benefits outweigh these investments.

By improving the accuracy and speed of medicinal plant disease classification—especially in commonly used plants like basil—the project can contribute to minimizing crop losses, reducing dependency on expert manual inspection, and enhancing the quality of herbal products in the market. This, in turn, can result in higher yields, better-quality raw materials for the pharmaceutical and herbal industry, and increased profits for farmers and manufacturers.

The system can be deployed using cost-effective open-source tools like Python, TensorFlow, Keras, and OpenCV. Moreover, the use of general-purpose computing infrastructure (such as standard laptops with GPU support) ensures affordability for small-scale farms and research centres.

A cost-benefit analysis reveals that the reduction in manual labor, minimized error in disease detection, and increased efficiency in quality control make this project financially viable and sustainable in the long term. Additionally, scalability to other medicinal plants further increases the return on investment.

The project demonstrates strong economic feasibility, especially when viewed through the lens of improved productivity, reduced losses, and enhanced quality standards in the medicinal plant supply chain.

**3.2.3 OPERATIONALFEASIBILITY**

The operational feasibility of the Classifier to Assess the Quality of Medicinal Plants Leaf project is determined by its practicality and ease of implementation in agricultural and herbal medicine environments.

Key considerations include the availability of medicinal plants Leaf image datasets, integration with existing agricultural workflows (such as nursery inspections, herbal processing units, or farm quality control), and the system’s ability to function in resource-constrained settings. The system is designed to be lightweight and user-friendly, requiring minimal technical expertise to operate—making it suitable even for rural farm operators or small-scale herbal businesses.

14

The project leverages widely used technologies such as Python, OpenCV, and CNN-based models, ensuring scalability and maintainability. Its web-based or mobile-compatible design allows flexibility in deployment across various platforms, including PCs, tablets, and smartphones.

Operational deployment does not require high-end infrastructure and can be integrated seamlessly into existing agricultural monitoring systems. Furthermore, no regulatory barriers are expected as the tool does not involve patient data or invasive methods.

By addressing these operational aspects such as ease of use, accessibility, scalability, and minimal training requirements the project is shown to be operationally viable and ready for practical deployment across herbal farms, agricultural extension centers, and research institutes.

**3.2.4 TIME FEASIBILITY**

The time feasibility of the “Design a Classifier to Assess the Quality of Medicinal Plants Using Learning Algorithms” project depends on several factors, including the scope of the project, availability of resources, and efficiency of project planning and execution.

Proper scheduling, phased implementation, and milestone tracking are essential to ensure that the project is completed within a realistic and achievable timeline. Key stages such as dataset collection, image preprocessing, model development, training, testing, web-based or app-based interface creation, and final deployment have been carefully planned.

The availability of a pre-curated dataset of basil plant images, along with the use of high-level libraries and frameworks like TensorFlow, Keras, and OpenCV, significantly accelerates the development cycle. Regular monitoring of progress and quick resolution of issues ensure the timely completion of each phase.

With well-defined goals, efficient use of open-source tools, and a focused scope, this project demonstrates strong time feasibility and is expected to be completed within the planned academic schedule.

**3.3 RISK ANALYSIS**

Risk analysis is a vital component of this project, as several potential challenges and uncertainties could affect its implementation and success.

First, data quality and availability present a medium-level risk. Although a dataset of basil plant images is available, ensuring diversity and accurate labeling is essential for effective model training.

Algorithm performance is considered a high-level risk, especially in real-world conditions where lighting, background noise, and image variations could affect prediction accuracy.

Interpretability and explainability of deep learning models present a medium-level risk, as their “black box” nature may make it difficult for non-technical users (like farmers or herbalists) to trust the system’s decisions.

Implementation challenges, such as deploying the system in low-resource environments or ensuring usability on mobile and web platforms, also pose a medium-level risk.

Although the system does not handle sensitive user data, ethical considerations like transparency and fairness in model predictions must be addressed—classified as a low to medium-level risk.

Resource constraints, including limited access to high-performance GPUs, budget restrictions, or limited technical support, represent another medium-level risk.

Lastly, the rapid evolution of AI technologies is a high-level risk, as new models or frameworks could quickly outperform current solutions, requiring ongoing learning and updates.

15

Proper risk mitigation strategies—such as data validation, model tuning, continuous testing, and modular system design—will help minimize these risks and support a robust, long-term solution.

|  |  |
| --- | --- |
| **Risk Factor** | **Probability (High/Medium/Low)** |
| Data Quality and Availability | Medium |
| Algorithm Performance | High |
| Interpretability and Explainability | Medium |
| Implementation Challenges | Medium |
| Ethical Considerations | Medium |
| Resource Constraints | Medium |
| Rapid Technological Advancements | High |

**Table 3.2: Risk Probability Matrix**

16

Chapter 4

**SOFTWARE REQUIREMENTS SPECIFICATION**

**4.1 INTRODUCTION**

The Software Requirement Specification (SRS) serves as a fundamental document in the development of the project. This document outlines the software requirements and specifications necessary to achieve the project's objectives. The SRS provides a comprehensive overview of the system's scope, functionalities, and constraints. It serves as a reference point for the project team, stakeholders, and developers, ensuring a common understanding of the software's purpose and expected outcomes. The SRS plays a crucial role in guiding the development process, facilitating effective communication, and ensuring that the final product meets the desired criteria. By clearly defining the software requirements, the SRS forms the foundation upon which the project implementation, testing, and evaluation phases are built.

**4.2 Project Scope**

* **Automated Medicinal Plant Health Assessment**: The developed system automates the process of assessing the health and quality of medicinal plants, particularly basil, by analysing leaf images using deep learning models.
* **Support for Farmers and Herbal Industries**: The system provides a valuable tool for farmers, herbal product manufacturers, and agricultural researchers by accurately detecting plant diseases and ensuring high-quality raw material.
* **Remote Plant Health Monitoring**: The project can be extended to support remote monitoring through mobile or web-based platforms, enabling users in rural or remote locations to assess plant health without the need for on-site experts.
* **Research and Agricultural Data Analysis**: The project contributes to research in the field of agricultural AI by offering a framework for automated medicinal plant classification, which can be used for further academic and commercial studies.

**4.3 External Interface Requirements**

#### ****4.3.1 User Interface****

The user interface is a crucial component of the system, offering an intuitive and user-friendly platform for interaction between users such as farmers, researchers, or herbal manufacturers and the medicinal plant quality assessment tool. It should allow users to upload or capture images of basil leaves, initiate analysis with a single click, and view clearly presented classification results (e.g., plant name, disease detected, health status, and medicinal usability). The interface should be designed to support users with varying levels of technical expertise and ensure a smooth and efficient workflow.

17

#### ****4.3.2 Hardware Interface****

The system should be compatible with basic hardware including standard desktop or laptop computers, as well as mobile devices with camera capabilities. It should also support image inputs from smartphones, USB webcams, or digital cameras used to capture basil plant images. No specialized imaging equipment is required, making it ideal for use in field conditions and low-resource environments.

#### ****4.3.3 Software Interface****

The software system is built using Python, with frameworks such as Flask for the web interface and TensorFlow/Keras for deep learning. It integrates modules for image preprocessing (OpenCV), model prediction, and result display. Python’s modular design ensures a clean separation between UI, backend logic, and model inference, simplifying future updates or enhancements.

#### ****4.3.4 Communication Interface****

For systems integrated with cloud or network-based setups, the communication interface should enable the secure upload of plant images and retrieval of classification results. While privacy concerns are minimal (since plant images do not involve sensitive user data), secure and reliable data transmission protocols (e.g., HTTPS or API calls) should still be considered. The system is designed to support both offline standalone use and online deployments for broader accessibility.

**4.4 NON-FUNCTIONAL REQUIREMENTS**

 **Performance:**

* The system should process basil leaf images and deliver classification results (e.g., plant health, disease name) with minimal latency to enable real-time usability.
* It must handle multiple image uploads efficiently, even during simultaneous usage by different users or institutions.

 **Accuracy and Reliability:**

* The classifier should achieve high accuracy in distinguishing between healthy and diseased basil leaves to ensure meaningful output.
* It must produce consistent and repeatable results for the same input image, thereby ensuring trustworthiness and dependability.

 **Scalability:**

* The system should be scalable to include other medicinal plants in the future and accommodate an expanding user base or dataset.
* It should be capable of scaling computationally, especially if deployed on the cloud or as a web/mobile app.

18

 **Security and Privacy:**

* While the system does not deal with sensitive human data, it should still ensure basic image security, especially for online deployments, using secure protocols and access controls.
* Privacy policies regarding image usage and model predictions should be clearly defined.

 **Usability and User Experience:**

* The interface should be clean, intuitive, and friendly for non-technical users like farmers or herbal practitioners.
* The system must provide helpful prompts, progress indicators, and visual feedback (e.g., highlighted disease area or tooltips) to assist users effectively.

 **Compatibility:**

* The system should run on a variety of hardware devices including desktops, laptops, and smartphones.
* It should support popular operating systems like Windows, Linux, Android, and web browsers like Chrome and Firefox.

 **Maintainability and Extensibility:**

* The codebase should be modular and documented to support quick bug fixes and enhancements.
* It should use Git for version control and adhere to standard Python development practices.

 **Ethical Considerations:**

* The project must promote fairness and avoid biases—such as those introduced by image background, lighting, or leaf condition—that could affect model predictions.
* Transparency should be ensured by offering simple explanations for predictions or confidence levels to help users understand the model’s output.

**4.5 OTHER RREQUIREMENT**

**4.5.1 Dataset Requirements**

* For our Medicinal plant classifier, primarily focusing on Basil, Leaf samples were collected and verified by plant experts to ensure high data accuracy and reliability.
* We have prepared dataset of basil image 666 in total distributed into labelled sets based on disease category with augmentation resulting in 2986 images.
* This dataset has 5 subfolders which contain images of healthy Basil leaves and other 4 folders are diseased basil leaves namely ‘Downy Mildew’, ‘Fusarium Wilt’, ‘Gray Mold’, ’Septoria Leaf Spot’.

19

**4.5.2 Hardware Requirement**

* Desktop or laptop computer with x86 compatible CPU,
* 50GB or more of free disk space,
* At least 8 GB main memory is capable of running the system.
* Optionally A GPU with CUDA support can significantly accelerate the training and inference speed.
* A powerful computer with more CPU cores and a GPU will speed up the diagnosis procedure significantly, while the diagnosis time on a typical laptop is also acceptable (less than 20 s per image).

**4.5.3 Software Requirements:**

* Operating system: Windows 7 or newer, 64-bit macOS 10.13+, or Linux, including Ubuntu, RedHat, CentOS 6+, and others
* Python 3.7 and above with libraries like Pandas, Numpy, scipy, keras models, pickle, OpenCV, functools, pyplot, tkinter, Tensorflow, sklearn etc.
* Editor: VS Code or any python compatible editor.

**4.6 ANALYSIS MODEL**

**4.6.1 DATA Flow Diagram**

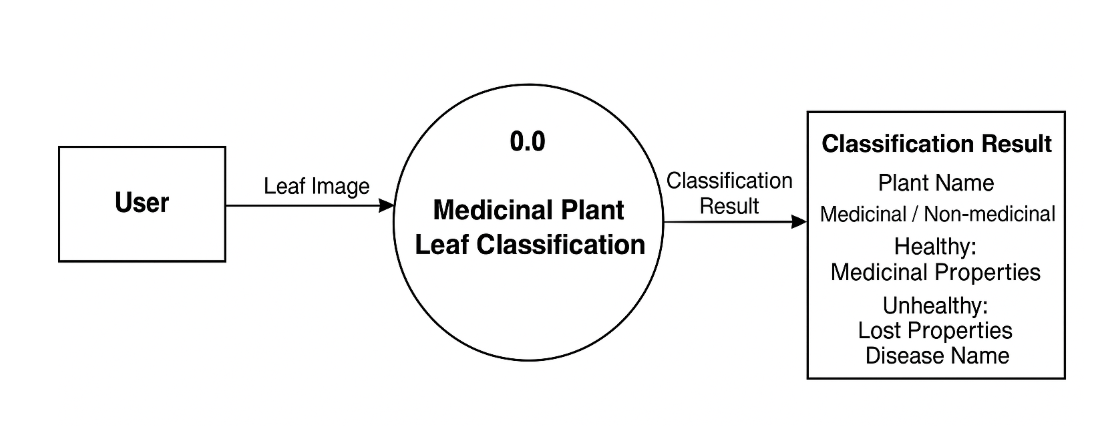
A data flow diagram (DFD) is a graphical representation of the "flow" of data through an information system, modeling its process aspects. A DFD is often used as a preliminary step to create an overview of the system without going into great detail, which can later be elaborated. DFDs can also be used for the visualization of data processing (structured design). A DFD shows what kind of information will be input to and output from the system, how the data will advance through the system, and where the data will be stored. A data flow diagram can dive into progressively more detail by using levels and layers, zeroing in on a particular piece. DFD levels are numbered 0, 1 or 2, and occasionally go to even Level 3 or beyond.

**Level 0:**

The necessary level of detail depends on the scope of what you are trying to accomplish. DFD Level 0 is also called a Context Diagram. It’s a basic overview of the whole system or process being analyzed or modeled.

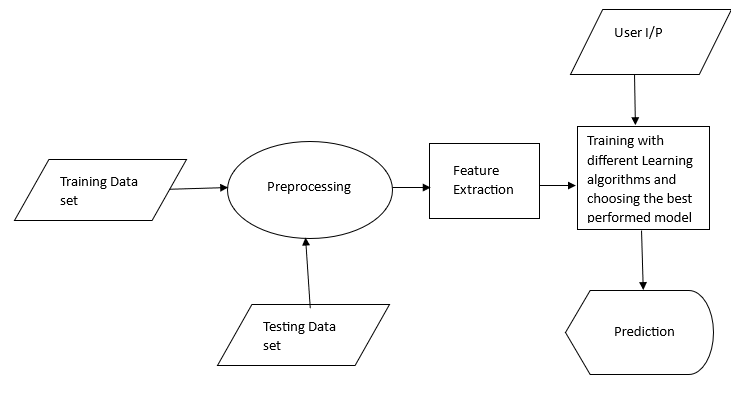
20

**Figure 4.1: Level 0 DFD**



Healthy/Unhealthy

**Level 1:**

****It’s designed to be an at-a-glance view, showing the system as a single high-level process, with its relationship to external entities.

**Figure** **4.1: Level 1 DFD**

21

**4.7 System Implementation Plan**

|  |  |  |  |
| --- | --- | --- | --- |
| **Phase** | **Task** | **Duration** | **Remark** |
| Phase -1 | Literature Survey | Aug - Sep | Done |
| Phase -2 | Study of Existing techniques for their strengths and weakness | Oct - Nov | Done |
| Phase -3 | Dataset selection and GUI design | Nov - Dec | Done |
| Phase -4 | Selection and Implementation of algorithms with appropriate techniques | Jan - Feb | Done |
| Phase -5 | Testing of hypothesis and performance analysis of system | Feb - March | Done |
| Phase -6 | Validation of proposed system. | March - Apr | Done |

**Table 4.1: System Implementation Plan**

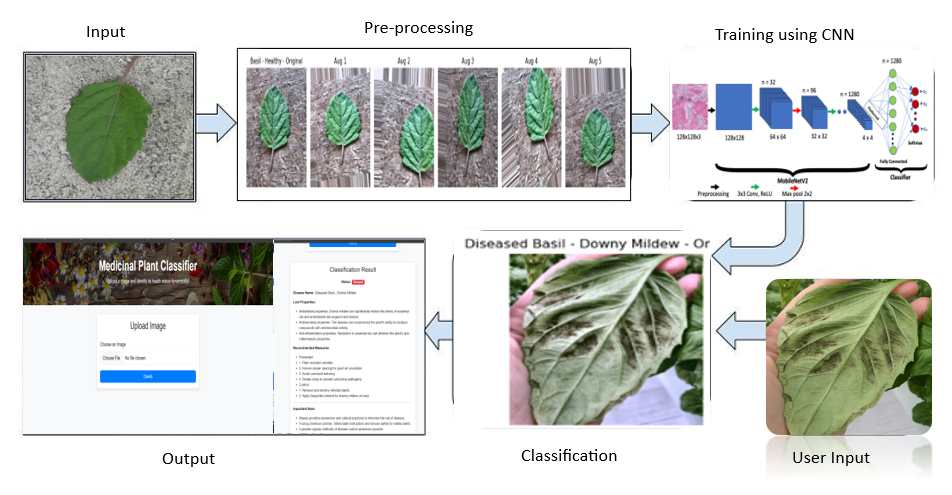
22

Chapter 5

**SYSTEM DESIGN**

**5.1 SYSTEM ARCHITECTURE**

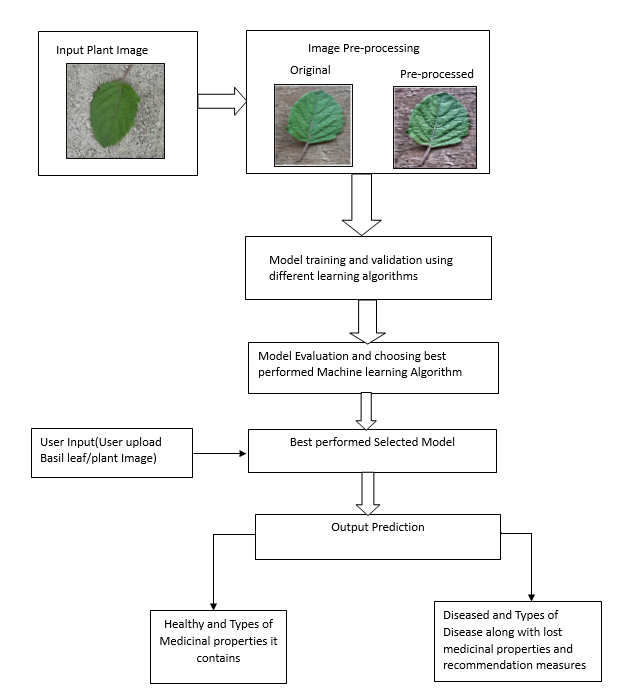
### To design an effective classifier for assessing the quality of medicinal plants, our methodology will follow a structured series of steps aimed at ensuring a comprehensive approach to data collection, feature selection, and model section, model development.

****The figure 1 below shows the system architecture of proposed system.

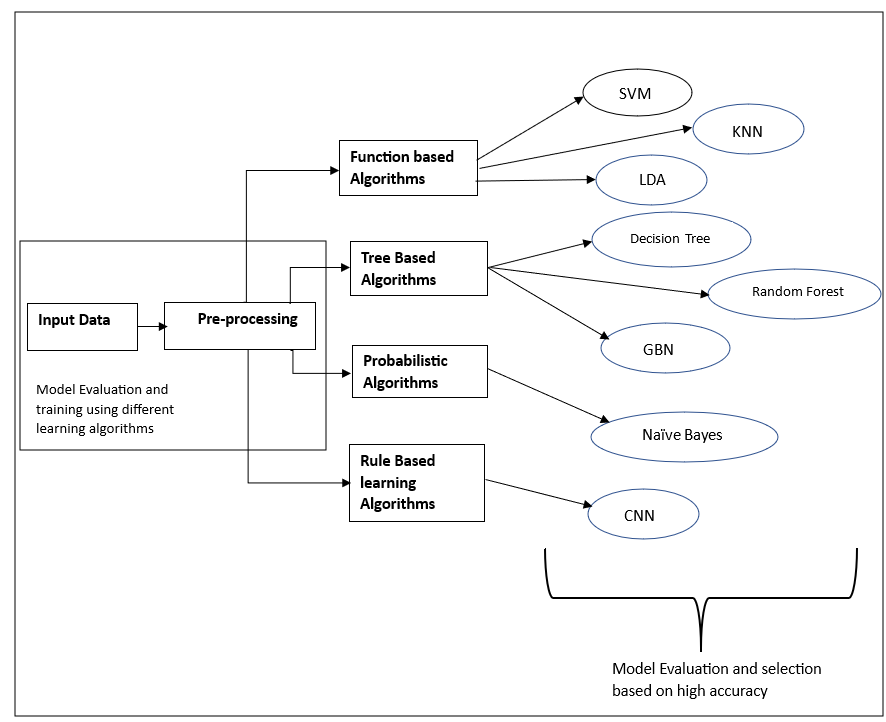
**Figure 5.1. System Architecture of Proposed classifier**

To Implement this System Architecture, our model follows many steps as presented below:

23

 **Figure 5.2: Working Methodology**

24

****

**Figure 5.3: Model Section Strategy**

**5.1.1 DATASET COLLECTION AND PREPRATION**

Our dataset consists of 666 basil leaf images, categorized into five classes: Healthy, Downy Mildew, Fusarium Wilt, Gray Mold, and Septoria Leaf Spot. The distribution includes 149 healthy images, 124 Downy Mildew, 212 Fusarium Wilt, 152 Gray Mold, and 29 Septoria Leaf Spot samples.

**Dataset samples:**

****Figure 2 shows healthy basil leaf samples. They do not have any disease. They are dark green and do not have any spots or discoloration.

**Figure 5.4 Healthy Basil image.**

25

Figure 3 shows Downy Mildew diseased basil leaves samples. These leaves show yellowing in areas restricted by major veins, with a fuzzy gray growth on the underside of leaves.

**Figure 5.5 Downy Mildew Disease**

Figure 4 shows Fusarium Wilt diseased images. Plants affected by this disease show stunted growth and dark streaks on the lower side of leaf and stem. Leaves often appear wilted even when soil moisture is adequate.

**Figure 5.6 Fusarium Wilt Disease**

Figure 5 shows Gray Mold affected basil leaves. This disease appears as Gray-brown fuzzy patches that spread rapidly in humid conditions.

**Figure 5.7 Gray Mold Disease**

****Septoria Leaf Spot diseased leaves have tiny dark spots with light centres, as shown in below Figure 6.

**Figure 5.8 Septoria Leaf Spot Disease.**

**5.1.2 Data Pre-Processing**

Preprocessing was a crucial phase to ensure uniformity and quality across all images. Several steps were performed as below:

* **Image Standardization**:

26

All images were resized to 224×224 pixels to safeguard consistent input dimensions for neural networks, reduce computational complexity, and eliminate scale variations across the dataset. This dimension was selected to align with common neural network architecture requirements.

**Figure 5.9: Resized Image**

* **Contrast Enhancement**:

Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied to accentuate morphological features. CLAHE processes the image in small regions, equalizes each region independently, and limits contrast enhancement to reduce noise amplification. An 8×8 tile grid with a clip limit of 2.0 was used, applying CLAHE to the luminance channel in LAB color space to reserve color information while enhancing structural details.

**Figure 5.10: Contrast Enhancement**

* **Noise Reduction :**

Gaussian blur filtering with a 5×5 kernel size was implemented to minimize noise while keeping relevant edges and structures. This technique effectively minimizes high-frequency noise elements that could interfere with classification.

Gaussian blur filter mathematically presented as:

Where,

* G(x,y) is the gaussian kernel value at point (x,y).
* is the standard deviation (controls the spread of the blur).

**Figure 5.11: Noise Reduction**

27

* **White Balance correction:**

A white balance algorithm was implemented to normalize color representation across images taken under different lighting conditions. This technique converts images to LAB color space, calculates average chrominance values, and adjusts channels based on pixel luminosity, effectively compensating for color temperature variations.

Mathematically presented as:

Where,

* A’, B’: Corrected chrominance values.
* , : Target values (usually 128 in OpenCV for neutral color).
* : Average values of the A and B channels.

**Figure 5.12 : White balance correction**

After applying all the above-mentioned techniques, we received final result as below for image pre-processing:

**Figure 5.13: Concluding results after applying all image pre-processing techniques**

**5.1.3 Dataset Split:**

We split the augmented dataset into three sets as training, validation, and testing:

* + Training set: 70% (464 images)
  + Validation set: 15% (100 images)
  + Testing set: 15% (102 images)

The split was performed using stratified sampling to maintain class distributions across all sets.

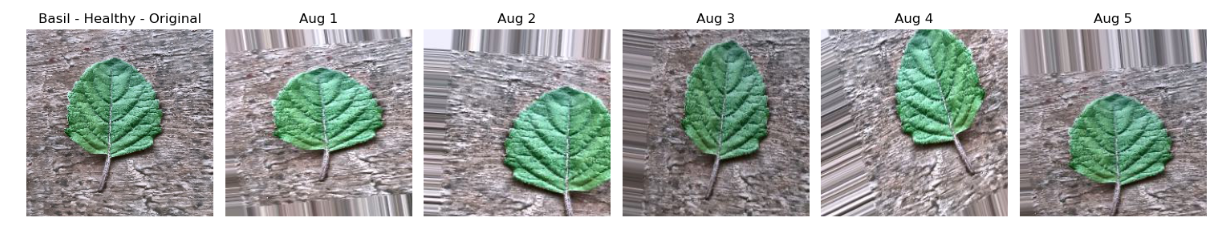
**5.1.4** **DATA AUGMENTATION**

To address class imbalance and improve generalization for training dataset, we applied multiple augmentation techniques:

* + Random rotation (±20°)
  + Horizontal and vertical flips.
  + Brightness/contrast adjustments (±10%)
  + Zoom variations (0.9-1.1×)

28

These techniques expanded our dataset from 666 to 2,986 images, with sufficient training samples for every disease category.



**Figure 5.14: Data Augmentation of sample images**

**5.1.5 FEATURE EXTRACTION AND PCA**

We extracted 120 features from 2,784 basil images, including 50 color features, 56 texture features, and 14 shape features. To improve efficiency, we applied Principal Component Analysis (PCA), reducing dimensions from 120 to 23 while preserving 95.42% of information. The principal components primarily captured texture homogeneity, color distributions, and shape characteristics. This approach enhanced computational performance and classification accuracy by focusing on the most distinctive features of basil leaves.

**5.1.6 MODEL SELECTION AND PERFORMANCE EVALUATION**

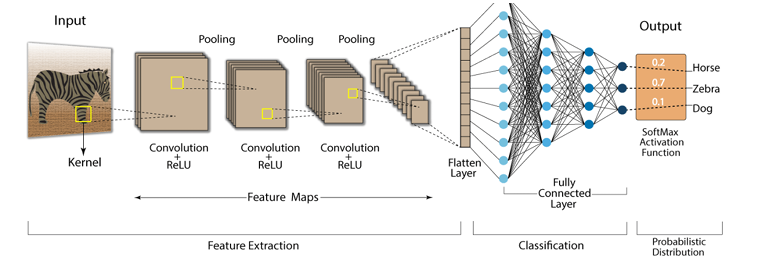
Here, we tested with different learning algorithms across different paradigms to classify basil images based on extracted features. Below we explain how algorithm works and their respective performance.

* **Function-Based Algorithms**
* Support Vector Machine (SVM) - 85.89% accuracy: SVM works by finding the hyperplane that maximizes the margin between classes in feature space. For non-linearly separable data, it uses kernel functions to transform the data in more dimensions where linear separation becomes possible. Our SVM implementation likely benefited from the PCA-transformed features, allowing effective separation of basil classes.
* K-Nearest Neighbors (KNN) - 73.09% accuracy: KNN classifies new samples based on the majority class among their k-nearest Neighbors in feature space. It uses distance metrics (typically Euclidean) to determine proximity. KNN is non-parametric and makes no conventions about data distribution, but develops computationally affluent with large datasets.
* Linear Discriminant Analysis (LDA) - 62.44% accuracy: A linear model assuming normal distribution and equal class covariance, which may not fit our data perfectly.

29

* **Tree-Based Algorithms:**
* Decision Tree – 62.2% accuracy: Decision Trees divide the dataset into subsets based on feature values, producing a tree-like model for decision-making. In this model, internal nodes are responsible for assessing features, branches represent the consequences of these evaluations, and leaf nodes denote the final class predictions. Despite their simplicity and interpretability, decision trees are susceptible to overfitting, which can decrease their overall predictive accuracy.
* Random Forest - 82.66% accuracy: Random forest is a learning algorithm that constructs a collection of decision trees, each trained on dissimilar subsets of the dataset created through random sampling with replacement. Additionally, it introduces randomness by selecting a set of features for each split, promoting variety among the trees. The concluding prediction is made by combination of all outputs of all trees, which helps reduce overfitting and increases model accuracy. This method is well-suited for enhancing generalization compared to using a single decision tree.
* Gradient Boosting (GBN) – 77.87% accuracy: Gradient Boosting builds a series of decision trees in a sequential manner, where each new tree is trained to correct the errors made by its predecessors. This iterative correction process helps the model focus on difficult examples, leading to better accuracy through adaptive learning and performance refinement.
* **Probabilistic Algorithms:**
* Gaussian Naive Bayes - 56.1% accuracy: This algorithm applies Bayes' theorem with an assumption that features follow a Gaussian distribution and are conditionally independent given the class. Its relatively poor performance suggests that either the independence assumption doesn't hold for our features or the feature distributions aren't Gaussian.
* **Rule-Based Learning Algorithms:**
* Convolutional Neural Network (CNN) - 89.22% accuracy: Our CNN implementation leverages the MobileNetV2 architecture, which is optimized for mobile and edge devices while maintaining high accuracy. The model employs depth-wise discrete convolutions that knowingly reduce computational complexity compared to standard convolutions. For our basil classification task, we utilized transfer learning by adopting pre-trained weights from ImageNet, freezing the base layers to preserve general feature extraction capabilities. We then modified the model with a classification head containing of global average pooling, a dense layer with 128 neurons, and dropout regularization to prevent overfitting. This architecture efficiently captured distinctive visual patterns in basil leaves—from basic textures to complex morphological features—resulting in superior classification performance while necessitating less parameters than traditional CNN architectures.

30

**Figure 5.15: Architecture of CNN**

**5.2 UML DIAGRAMS**

* **Class Diagram**

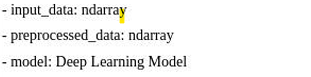
A class diagram is a type of static structure diagram in UML (Unified Modeling

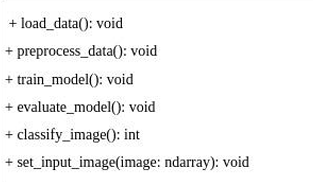
Language) that represents the structure and relationships of classes in a system. It

provides a visual representation of the classes, their attributes, methods, and

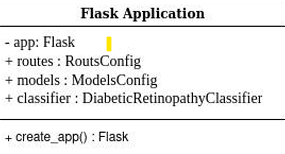
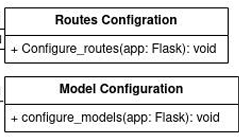
associations, as well as the inheritance and dependency relationships between classes.

Medicinal Plants Leaf Classification





0.1



1

0.1

1

1

0.1

**Figure 5.16: Class Diagram**

31

* **Sequence Diagram**

A sequence diagram is a type of interaction diagram in UML (Unified Modeling

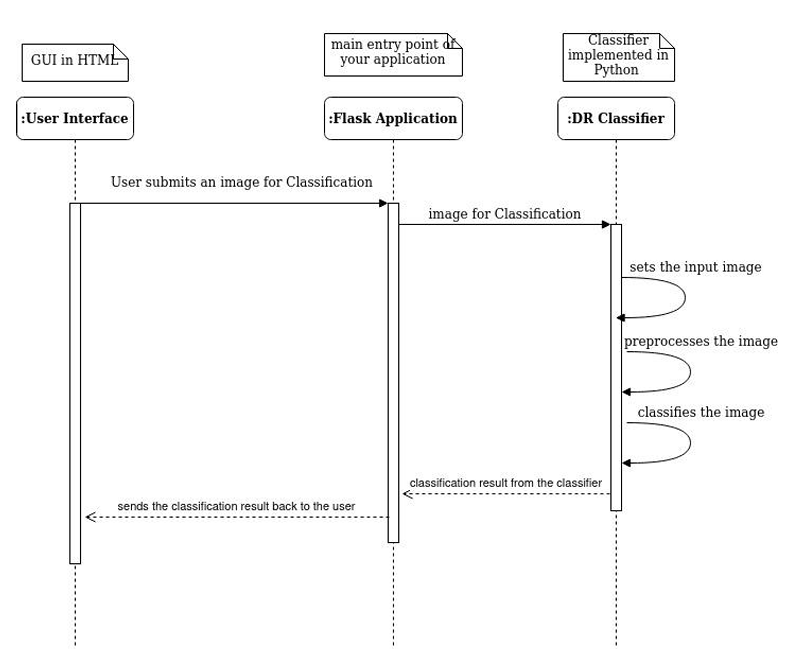
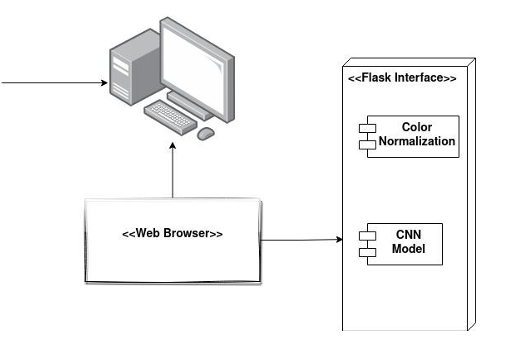
 Language) that illustrates the interactions between objects or components in a system over time. It focuses on the sequence of messages exchanged between the objects to accomplish a specific behaviour or scenario. Sequence diagrams are useful for visualizing the flow of control and data between different parts of a system. They depict the order of method invocations, along with the objects or actors participating in the interactions.

Figure 5.17: Sequence Diagram

32

* **Deployment Diagram**

Deployment diagrams are useful for understanding the hardware infrastructure required for running a software system, including the placement of components on nodes and their interconnections. They help visualize the physical aspects of a system deployment, including the allocation of software components to hardware resources and the relationships between them.



Integration of selected model

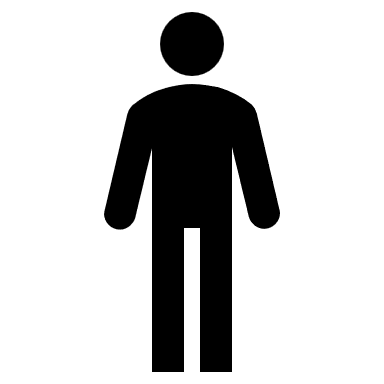


Figure 5.18: Deployment Diagram

33

**Chapter 6**

**TECHNICAL SPECIFICATIONS**

**6.1 ADVANTAGES**

* **Accurate Medicinal Plant Classification:**

This project leverages deep learning techniques, specifically Convolutional Neural Networks (CNNs), to accurately classify basil plant leaves into healthy or diseased categories (such as Downy Mildew, Fusarium Wilt, Gray Mold, and Septoria Leaf Spot). This ensures high classification accuracy, reducing reliance on manual inspection and expert consultation, and thereby improving the reliability of medicinal plant assessment.

* **Standardized Image Analysis through Preprocessing :**

Although color normalization is not used, the system incorporates essential preprocessing steps such as **image resizing, noise reduction, and normalization** to maintain uniformity across all input images. This improves the consistency of results and enhances the CNN model’s ability to correctly interpret leaf characteristics, regardless of imaging conditions.

* **Scalability and Real-Time Processing :**

Once trained, the CNN model is capable of analyzing hundreds of plant images efficiently. The system is designed to be scalable, making it suitable for implementation in agricultural research centers, herbal farms, and mobile applications. Its fast response time supports real-time classification, enabling immediate insights into plant health.

* **Early Detection for Crop and Quality Protection :**

Diseases in medicinal plants can reduce or eliminate their healing properties. This system enables **early detection** of such diseases before they become widespread, allowing farmers or pharmaceutical companies to take corrective measures, **protect crop value**, and maintain quality standards in herbal medicine production.

* **Objective, Consistent, and User-Friendly Results:**

Unlike manual inspections, which can vary between individuals, this AI-driven system provides consistent, unbiased, and repeatable results. Its user-friendly interface ensures that even individuals without technical expertise can easily upload images, view classification results, and access recommendations.

Overall, the system offers significant advantages such as accurate plant health classification, consistent output through standardized preprocessing, scalability for wide-scale deployment, efficiency in real-time operation, and early disease detection to maintain medicinal value. These benefits collectively contribute to increased productivity, improved disease management, and enhanced quality assurance in the medicinal plant supply chain.

34

**6.2 APPLICATIONS**

* **Automated Medicinal Plant Quality Assessment:**

The developed deep learning-based classification system can be used to automate the assessment of medicinal plant leaves, specifically basil. It enables users to quickly and accurately identify whether a plant is healthy or diseased, improving the efficiency and reliability of quality control processes in herbal medicine production.

* **Early Disease Detection in Agricultural Settings:**

The system supports early detection of plant diseases, allowing timely interventions that can prevent widespread damage. By identifying symptoms of common diseases such as Downy Mildew, Fusarium Wilt, and Gray Mold at an early stage, farmers can apply targeted treatments and preserve the medicinal integrity of their crops.

* **Integration with Mobile Apps for Field Use:**

The classifier can be integrated into mobile applications, enabling on-the-spot analysis of plant leaves using smartphone cameras. This application is particularly valuable for rural farmers, herbalists, and agricultural inspectors who require real-time feedback without needing advanced lab equipment.

* **Support for Herbal Industries and Pharmacies:**

Herbal product manufacturers and Ayurvedic pharmacies can use this system to ensure the medicinal plants used in their formulations are of verified quality. The classifier helps maintain high standards by filtering out unhealthy or compromised plants from the production pipeline.

* **Research and Agricultural Data Analysis:**

The implementation of this system can contribute to academic and industrial research in plant health monitoring, disease prediction, and agricultural AI. By analyzing classification patterns and outcomes, researchers can gain deeper insights into the impact of disease on medicinal properties and work toward improving disease-resistant plant varieties.

35

**Chapter 7**

**TESTING**

**7.1 System Testing:**

System Testing is a type of software testing that is performed on a complete integrated system to evaluate the compliance of the system with the corresponding requirements. In system testing, integration testing passed components are taken as input. The goal of integration testing is to detect any irregularity between the units that are integrated together. System testing detects defects within both the integrated units and the whole system. The result of system testing is the observed behaviour of a component or a system when it is tested.

**7.2 Integration Testing:**

Integration testing is the second level of the software testing process after unit testing. In this testing, units or individual components of the software are tested in a group. The focus of the integration testing level is to expose defects at the time of interaction between integrated components or units.

**7.3 Unit Testing:**

Unit testing uses modules for testing purposes, and these modules are combined and tested in integration testing. The Software is developed with a number of software modules that are coded by different coders or programmers. The goal of integration testing is to check the correctness of communication among all the modules.

|  |  |
| --- | --- |
| Test Case number | TC\_01 |
| Module Under Test | Website Directory Structure |
| Description | While using the Flask, we should make sure that the directory structure is well assigned to display images properly |
| Output | If the directory structure is correct the Input Image section, Classify Button will be displayed on website |
| Remarks | Test Successful |

Table 7.1: Test Case – 01

|  |  |
| --- | --- |
| Test Case number | TC\_02 |
| Module Under Test | Input Results Selection and classification |
| Description | User should be able to choose Input Image from the gallery and upload it. User should be able to click on classify button to get predictions |
| Output | After clicking on choose file button, user was able to choose the file and able to upload. User was able to click on classify button. |
| Remarks | Test Successful |

Table 7.2: Test Case 02

36

|  |  |
| --- | --- |
| Test Case number | TC\_03 |
| Module Under Test | Predictions for Healthy Medicinal Plants Leaf Image |
| Description | User should upload Healthy leaf image of medicinal plant and output should predict as Healthy along with Benefits. |
| Output | Classifier should able to predict output correctly as Healthy for Healthy Leaf Image |
| Remarks | Test Successful |

Table 7.3: Test Case – 03

|  |  |
| --- | --- |
| Test Case number | TC\_04 |
| Module Under Test | Predictions for Diseased Medicinal Plants Leaf Image |
| Description | User should upload Diseased leaf image of medicinal plant and output should predict as Diseased along with Disease name, lost properties and Recommendation Measures. |
| Output | Classifier should able to predict output correctly. |
| Remarks | Test Successful |

Table 7.4: Test Case - 04

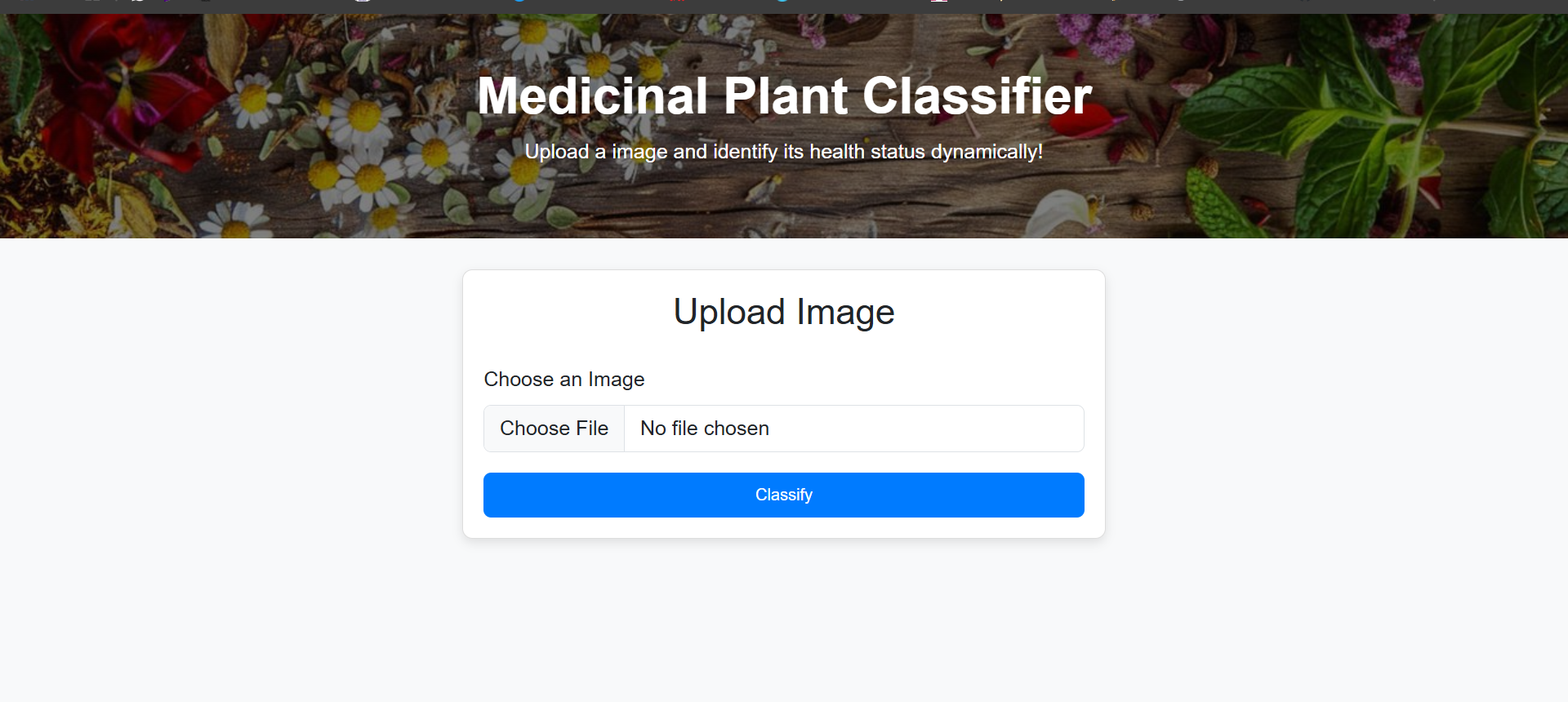
37

Chapter 8

**RESULTS AND DISSCUSSIONS**

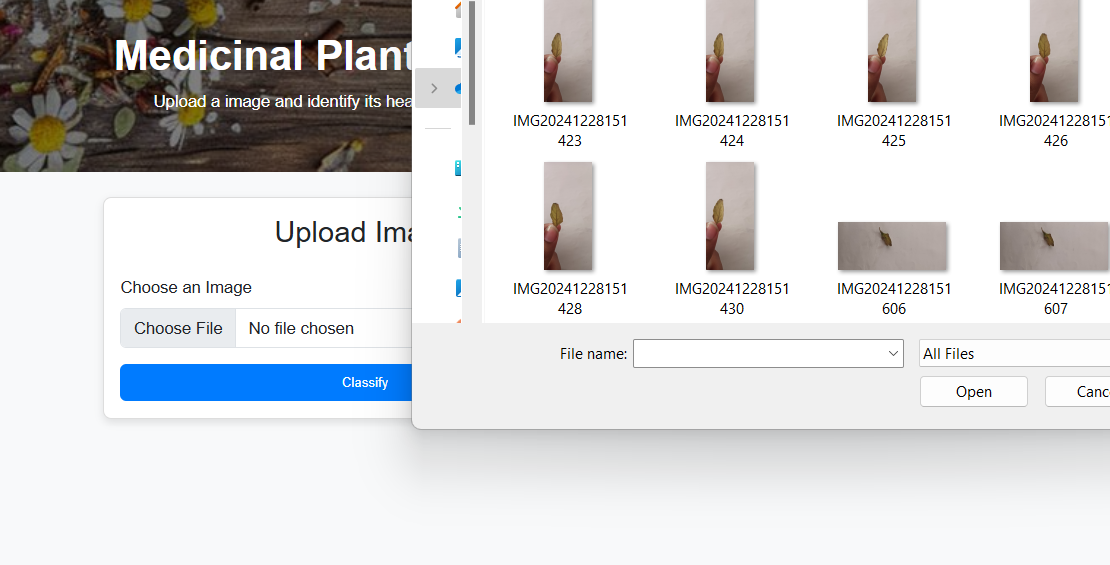
**8.1 RESULTS**

* **Home Page**



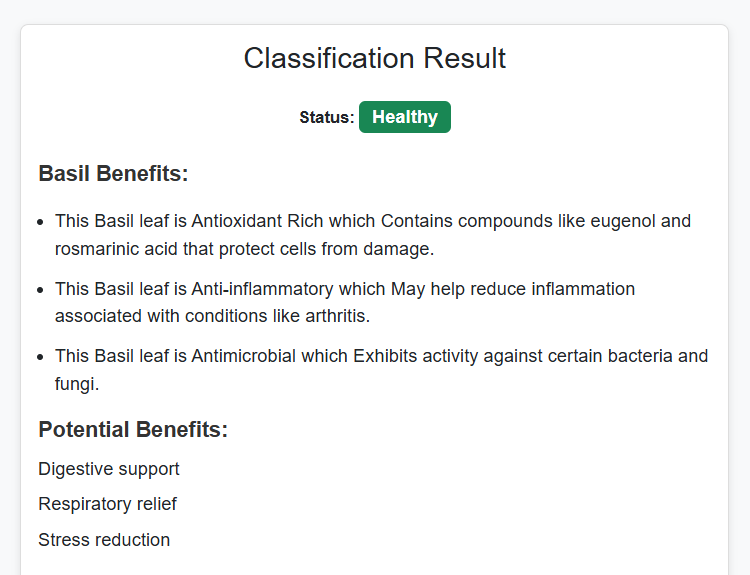
**Figure 8.1 : Home Page**

* **Input Selection**



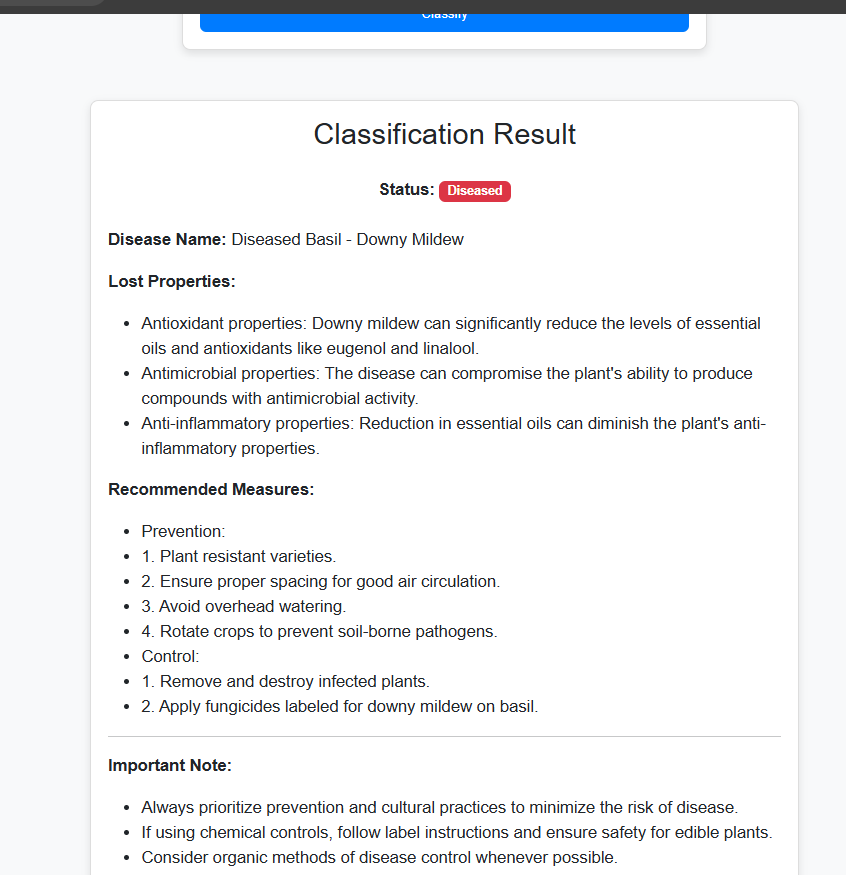
**Figure 8.2 : Input Selection**

38

* **Output Prediction if uploaded Medicinal Plants Leaf(Basil) is Healthy :**

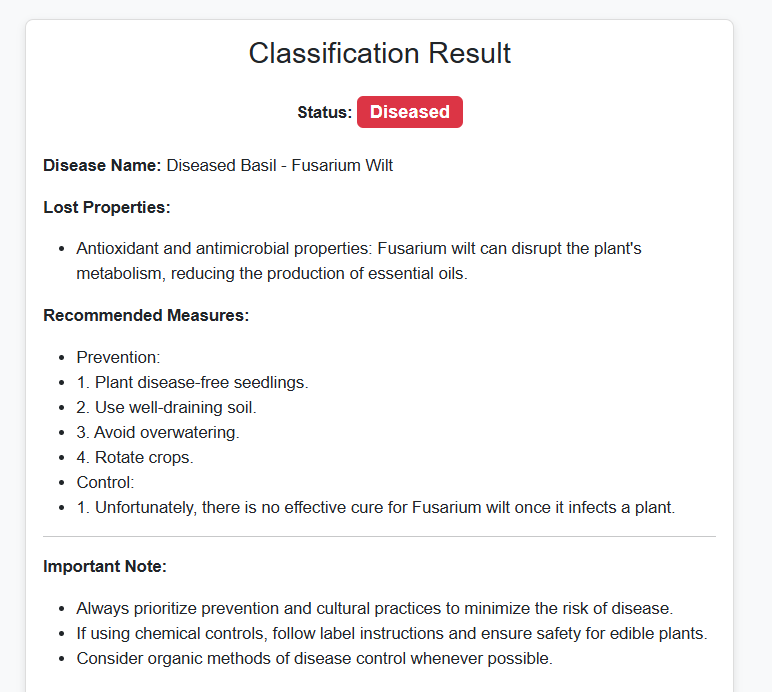
**Figure 8.3: Prediction as Healthy**

* **Output Prediction if Basil has Downey Mildew Disease**

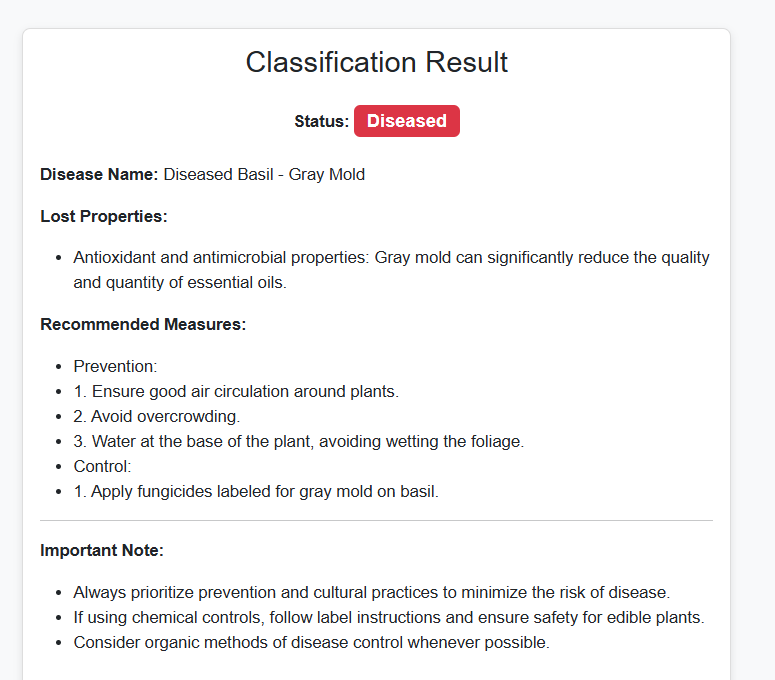


**Figure 8.4 : Prediction as diseased basil -1**

39

* **Output Prediction if Basil has Fusarium Wilt Disease:**

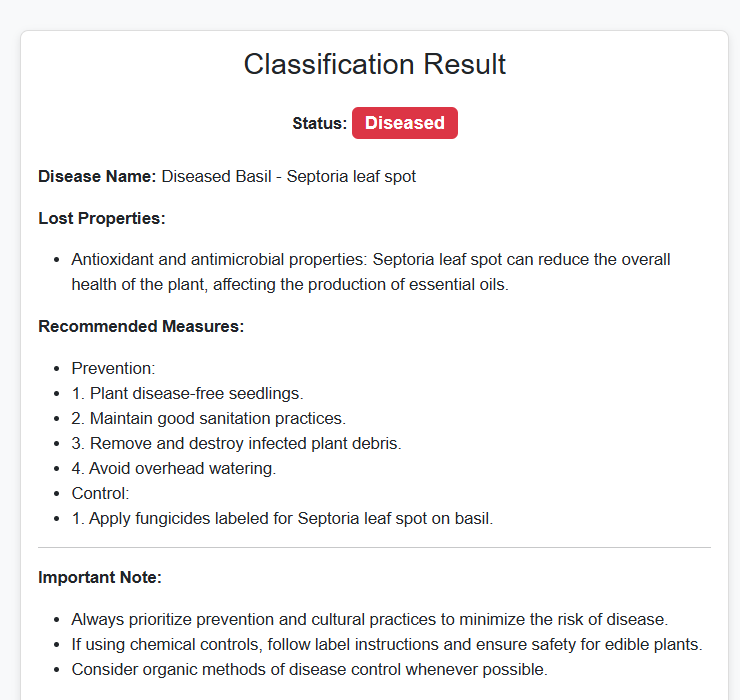
**Figure 8.5: Prediction as diseased basil -2**

* **Output Prediction if Basil has Gray Mold Disease:**

**Figure 8.6: Prediction as diseased basil -3**

40

* **Output Prediction if Basil has Septoria Leaf Spot Disease:**



**Figure 8.7: Prediction as diseased basil -4**

41

**8.2 Discussions:**

Our comprehensive evaluation of various machine learning paradigms reveals distinct performance patterns that merit further discussion. When comparing algorithms across different classification approaches, we observed notable variations in accuracy that provide insights into the nature of basil classification.

**5.1. Function-Based Algorithms**

Among function-based algorithms, Support Vector Machine (SVM) demonstrated remarkable performance with 85.89% accuracy, significantly outperforming other algorithms in this category. This superior performance can be attributed to SVM's ability to effectively find optimal hyperplanes in the high-dimensional feature space created by our PCA transformation. The algorithm's capacity to handle complex non-linear relationships using kernel functions made it particularly well-suited for distinguishing subtle differences in basil leaf features. In contrast, K-Nearest Neighbors (KNN) achieved moderate performance (73.09%), while Linear Discriminant Analysis (LDA) showed limited effectiveness (62.44%), suggesting that the relationships between basil classes in our feature space are not optimally separated by linear transformations.

**5.2. Tree-Based Algorithms**

Within the tree-based models, Random Forest showed strong performance (82.66% accuracy), demonstrating the value of ensemble learning. The model’s strategy of building multiple trees on bootstrapped subsets, along with random feature selection, effectively reduced variance and prevented overfitting. On the additional hand, the single Decision Tree model had a lower accuracy of 62.2%, underlining its vulnerability to overfitting and poor generalization. Gradient Boosting (GBN), which refines predictions through sequential learning, achieved 77.87% accuracy, positioning itself between the two in terms of performance.

**5.3. Probabilistic Algorithms**

**Gaussian Naive Bayes** was the least effective, achieving only **56.1% accuracy**. Its fundamental assumption—that features are statistically independent and normally distributed within each class—likely limited its ability to model the complex correlations present in the image features extracted from basil leaves.

**5.4. Rule-Based Learning Algorithms**

**CNN**, utilizing the MobileNetV2 backbone, achieved the **highest overall accuracy of 89.22%**. The success of this model can be attributed to its ability to learn and represent hierarchical visual features directly from image data, without the need for manual feature engineering. Transfer learning allowed the model to benefit from pre-trained knowledge, while the customized classification layers helped tailor it to our specific dataset. Its superior performance confirms the potential of deep learning in plant disease detection and quality assessment.

**5.5. Cross-Paradigm Comparison**

The accuracy ranking observed—**CNN > SVM > Random Forest > GBN > KNN > LDA > Decision Tree > Naive Bayes**—highlights the importance of non-linear modelling and ensemble techniques in capturing subtle distinctions in image-based classification tasks. Deep learning and advanced function-based methods outperformed others, suggesting that models capable of capturing complex relations are better suitable for this problem. Ensemble methods also proved effective, offering a balance between model complexity and generalization.

42

These outcomes suggest that, depending on the computational constraints, **CNN should be the first choice** for high-performance classification. However, **SVM and Random Forest** can serve as reliable alternatives when computational or data resources are limited.

43

Chapter 9

**CONCLUSION & FUTURE WORK**

* **Conclusion :**

This report presented a classification system to assess medicinal leaf health using various machine learning models. Among all approaches, CNN achieved the highest accuracy of 89.22%, followed by SVM and Random Forest. The use of PCA and robust pre-processing techniques contributed significantly to model performance. These results demonstrate the potential of AI-based systems in supporting medicinal plant quality evaluation. Future work can focus on expanding the dataset, improving real-time detection, and applying this method to other plant species.

* **Future Work:**
* We will work on the implementation the same classifier for other Medicinal Plants Leaf.
* As currently it’s a web-based application, we can work on developing Mobile based application for our classifier.

44

**REFERENCES**

[1] T Shen, Hong yu: Assessing the impacts of climate change and habitat suitability on the distribution and quality of medicinal plant using multiple information integration: Take Gentiana rigescens as an example. 10.1016/j.ecolind.2021.107376

[2] Min He, Yu Zhou: How to identify “Material basis-Quality markers” more accurately in Chinese herbal medicines from modern chromatography-mass spectrometry data-sets: Opportunities and challenges of chemometric tools. 10.1016/j.chmed.2020.05.006

[3] Hussain, Hongbin: Classical and Emerging Non-destructive Technologies for Safety and Quality Evaluation of Cereals: A Review of Recent Applications. 10.1016/j.tifs.2019.07.018

[4] Nikitha Modupalli, Mohan Naik: Emerging non-destructive methods for quality and safety monitoring of spices. 10.1016/j.tifs.2020.12.021

[5] J. Wang, Q. Chen, T. Belwal, et al.: Insights into chemometric algorithms for quality attributes and hazards detection in foodstuffs using Raman/surface enhanced Raman spectroscopy. 10.1111/1541-4337.12741

[6] P. Mishra, J.M. Roger, D.N. Rutledge, et al.: SPORT pre-processing can improve near-infrared quality prediction models for fresh fruits and agro-materials. 10.1016/j.postharvbio.2020.111271

[7] Y. Li, Y. Shen, C. Yao, et al.: Quality assessment of herbal medicines based on chemical fingerprints combined with chemometrics approach: A review. 10.1016/j.jpba.2020.113215

[8] L. Qi, F. Zhong, Y. Chen, et al.: An integrated spectroscopic strategy to trace the geographical origins of emblic medicines: Application for the quality assessment of natural medicines. 10.1016/j.jpha.2019.12.004

[9] K. Tahri, C. Tiebe, N. El Bari, et al.: Geographical provenience differentiation and adulteration detection of cumin by means of electronic sensing systems and SPME-GC-MS in combination with different chemometric approaches . 10.1039/C6AY01906D

[10] C. Pérez-Ràfols, N. Serrano, J.M. Díaz-Cruz: Authentication of soothing herbs by UV-vis spectroscopic and chromatographic data fusion strategy. 10.1016/j.chemolab.2023.104783

[11] C. Liu, Z. Zuo, F. Xu, et al.: Authentication of Herbal Medicines Based on Modern Analytical Technology Combined with Chemometrics Approach: A Review. 10.1080/10408347.2021.2023460

[12] Silky Sachar & Anuj Kumar: Deep ensemble learning for automatic medicinal leaf identification. International Journal of Information Technology . 2022, 14:3089-3097. 10.1007/s41870-022-01055-z

45

[13] S. ROOPASHREE,J. ANITHA3, DeepHerb: A Vision Based System for Medicinal Plants Using Xception Features. 2021, 10.1109/ACCESS.2021.3116207

[14] FRIMPONG TWUM , YAW MARFO MISSAH , STEPHEN OPOKU OPPONG: Textural Analysis for Medicinal Plants Identification Using Log Gabor Filters. 2022, 10.1109/ACCESS.2022.3196788

[15] Jibi G. Thanikkal, Ashwani Kumar Dubey , and M. T. Thomas: Unique Shape Descriptor Algorithm for Medicinal Plant Identification (SDAMPI) With Abridged Image Database. 2020, 10.1109/JSEN.2020.3002909

[16] Congcong Wang, Xiaobo Zhang, Tingting Shi, et al.: Classification of Medicinal Plants Astragalus Mongholicus Bunge and Sophora Flavescens Aiton Using GaoFen-6 and MultitemporalSentinel-2 Data .2022, 10.1109/LGRS.2021.3120125

[17] Kalananthni Pushpanathan, Marsyita Hanafi, Syamsiah Mashohor: Machine learning in medicinal plants recognition. 2021, 10.1007/s10462-020-09847-0

[18] Jing Wei Tan , Siow-Wee Chang , Sameem Abdul-Kareem, et al.: Deep Learning for Plant Species Classification Using Leaf Vein Morphometric. 2020, 10.1109/TCBB.2018.2848653

[19] L. C. De: Bio-Diversity and Conservation of Medicinal and Aromatic Plants. 10.15406/apar.2016.05.00186

[20] Jana Wäldchen · Patrick Mäder: Plant Species Identification Using Computer Vision Techniques. 10.1007/s11831-016-9206-z

[21] Umair Ahmad , Sidra Ashiq, Gran Badshah, et al.: Feature Extraction of Plant Leaf Using Deep Learning. 10.1155/2022/6976112

[22] Esther Salmerón-Manzano, Jose Antonio Garrido-Cardenas, Francisco Manzano-Agugliaro: Worldwide Research Trends on Medicinal Plants. 10.3390/ijerph17103376

[23] Halimatu Sadiyah: Convolution neural network in precision agriculture for plant image recognition and classification. 2017, 10.1109/INTECH.2017.8102436

**APPENDIX A – PUBLICATIONS**

|  |  |
| --- | --- |
| **Sr. No.** | **Particulars** |
| 1 | Paper Publication - Survey Paper |
| 2 | Paper Publication - Implementation Paper |



46

